

SCIENTIFIC REPORTS

OPEN

A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent

Bobby R. Innes^{ID} & Thomas U. Otto^{ID}

Multisensory signals allow faster responses than the unisensory components. While this redundant signals effect (RSE) has been studied widely with diverse signals, no modelling approach explored the RSE systematically across studies. For a comparative analysis, here, we propose three steps: The first quantifies the RSE compared to a simple, parameter-free race model. The second quantifies processing interactions beyond the race mechanism: history effects and so-called violations of Miller's bound. The third models the RSE on the level of response time distributions using a context-variant race model with two free parameters that account for the interactions. Mimicking the diversity of studies, we tested different audio-visual signals that target the interactions using a 2×2 design. We show that the simple race model provides overall a strong prediction of the RSE. Regarding interactions, we found that history effects do not depend on low-level feature repetition. Furthermore, violations of Miller's bound seem linked to transient signal onsets. Critically, the latter dissociates from the RSE, demonstrating that multisensory interactions and multisensory benefits are not equivalent. Overall, we argue that our approach, as a blueprint, provides both a general framework and the precision needed to understand the RSE when studied across diverse signals and participant groups.

Everyday life requires processing of multiple signals coming from different sensory modalities like audition and vision. An incoming phone call, for example, can be signalled either in one modality only (e.g. a ringtone as a unisensory signal) or in multiple modalities simultaneously (e.g. a ringtone and a flashing screen as redundant signals). As observed in the redundant signal paradigm (Fig. 1a), the benefit of redundant signals is that individuals respond faster compared to the unisensory components, which is the *redundant signals effect* (RSE) as described in classic experiments^{1–3}.

Given the broad scope of multisensory benefits, the RSE has been investigated for over 100 years with different research questions and diverse signal sets. Basically all sensory modalities have been tested, including audio-tactile⁴, visual-tactile⁵, and gustatory-olfactory pairings⁶. The largest body of research has focused on audio-visual pairings, but even within this domain, stimuli have varied considerably across studies. Some have used abstract signals, such as pure tones or noise sounds that were paired with flashes, letters, or simple geometric shapes^{7–9}. Others have used real-world signals, such as animal sounds and images¹⁰ or signals with emotional content¹¹. In most experiments, signals have had sudden-onsets and were presented briefly with little variation across trials, but the RSE was also observed when signals were randomly generated on each trial and presented within a continuous audio-visual background^{12,13}. While the basic RSE has reliably replicated, the diverse studies differ in several aspects including overall mean RTs and the exact size of the RSE. Due to these differences, a recent review argued that a common modelling framework for the RSE is unlikely to be found¹⁴. In this view, a comparative approach between studies is difficult and a complete understanding of the RSE is lacking.

To develop a common framework, it is a good starting point to build upon models of unisensory decision-making, which typically assume some key processing steps e.g.^{15–17}. Sensory evidence for a signal is available from the environment, and is sampled by neurons with corresponding receptive fields. Due to fluctuations in neuronal activity, the evidence is subject to *noise*. To minimise the influence of noise on signal detection,

School of Psychology & Neuroscience, St. Mary's Quad, South Street, St. Andrews, KY16 9JP, United Kingdom. Correspondence and requests for materials should be addressed to B.R.I. (email: bobby.richard.innes@gmail.com) or T.U.O. (email: to7@st-andrews.ac.uk)

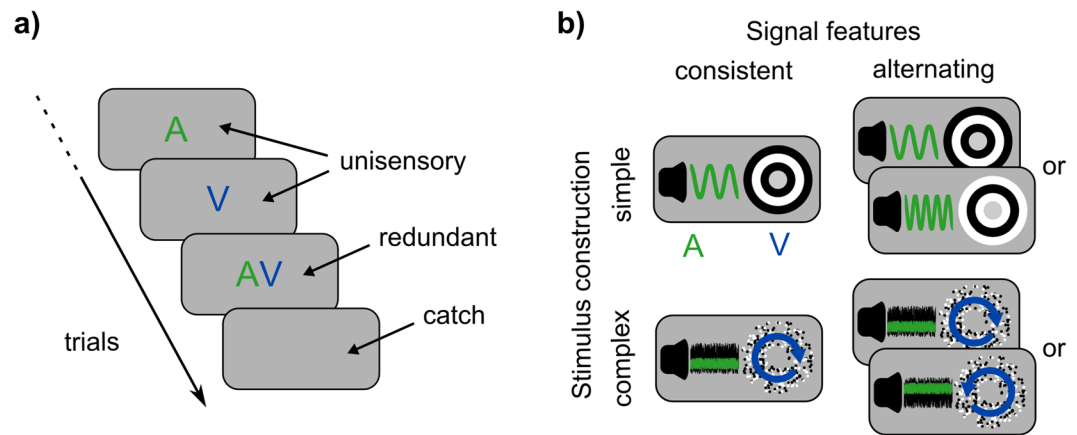


Figure 1. Experimental design. (a) Redundant signals paradigm. Participants respond to trials with auditory (A), visual (V), and combined signals (AV), but not on catch trials. Signals were presented in a random order within a block of trials. (b) Following a 2×2 design, we tested four signal sets in separate blocks. Stimulus construction was either simple (e.g. a pure tone, in audition) or complex (e.g. a noise tone). The sequence of signal features was either consistent (e.g. one frequency) or alternating (e.g. randomly one of two frequencies).

the evidence is accumulated over time until a threshold level is reached and a response is triggered. To extend the framework to the RSE, some sort of *combination rule* is needed that specifies how evidence from two senses is processed to trigger a single response. Broadly, there are two main model architectures with fundamentally different combination rules. Firstly, so-called *race models* suggest that evidence for each signal is accumulated separately by parallel decision units (e.g. one for audition and one for vision) and that the first unit to reach its threshold triggers a response. Secondly, *pooling models* propose that evidence for both signals is summed in a single decision unit and that a response is triggered when the combined evidence reaches a threshold level. Both architectures could in principle explain the RSE.

The race model architecture was first proposed by Raab¹⁸. The idea is that selecting the faster of two stochastic decision processes results on average in faster and less variable RTs with redundant compared to unisensory signals. Thus, race models explain the RSE mainly by statistical facilitation. Miller⁸, however, argued that if race models were correct, an upper bound for the RSE can be computed on the level of RT distributions, which is named Miller's bound. As this bound is typically violated in experiments, Miller⁸ and many follow-up studies concluded that race models cannot explain the RSE, which has led to the dominance of pooling models. However, to the best of our knowledge and in agreement with the recent review¹⁴, no pooling model has been shown to account for the RSE across studies. Hence, the lack of a common framework is in essence a failure of the dominating class of pooling models. The possibility that race models provide a common framework remained unexplored due to the general rejection.

This omission becomes critical as the general rejection of race models is not a valid argument¹⁹. The issue is the untested and mostly neglected *context invariance* assumption, which states that the processing of one signal is not affected by another^{20,21}. As Miller's bound implicitly assumes context invariance, a violation of the bound does not necessarily show that the basic race model architecture is wrong. Alternatively, the context invariance assumption may be wrong, which would show that some sort of interaction has taken place between the parallel decision units. If the context invariance assumption is dropped, we demonstrated that context variant race models can explain the RSE, including violations of Miller's bound^{12,13}. These considerations show that the parallel architecture of race models could in fact provide a common framework for RSE studies.

To study the RSE with diverse signal sets, here, we develop a comparative approach based on race models. In a first of three steps, we test the ability of the framework to predict multisensory benefits. Following Raab's model¹⁸, directional predictions of the RSE for different sets of signals can be readily made according to the *principles of multisensory behaviour*¹². Firstly, the *principle of equal effectiveness* states that benefits should increase when unisensory RTs become more similar. Secondly, the *variability rule* states that the variability of RTs is the key driving force of multisensory benefits, which should increase when unisensory RTs become more variable. In addition, following Raab's model¹⁸, a parameter-free prediction can be made of the exact size of the RSE based on the unisensory RT distributions and *probability summation*. The first step thus provides a simple tool to analyse the RSE if diverse signal sets yield different RTs.

The second step quantifies two processing interactions with multisensory signals that are not covered by Raab's model. The first interaction concerns *history effects* that describe the influence of previous trials on current-trial processing^{4,8,13}. Typically, RTs to unisensory signals are faster if a modality is repeated (e.g. audition following audition) compared to a switch (e.g. audition following vision). This interaction challenges the *statistical independence* assumption that is made by the parameter-free prediction using probability summation¹⁹. The second interaction is observed by violations of Miller's bound, which shows that processing of one signal must affect processing of the other signal in some way if the parallel processing architecture is correct. This second interaction challenges the *context invariance* assumption made by the parameter-free prediction based on probability

summation¹⁹. By quantifying the two interactions, it becomes clear that Raab's basic model needs to be extended. Moreover, the interactions need to be critically considered to understand how multisensory interactions are eventually changed with diverse signals.

The third step is to account for the RSE on the level of entire RT distributions with a modelling approach. Specifically, we use here the context variant race model that includes two free parameters to account for the interactions^{12,13,19}. Regarding history effects, the model includes the correlation Rho , which is used in the probability summation rule given that the assumption of statistical independence does not hold. Regarding violations of Miller's bound, the model includes the noise Eta , which models increased noise within the accumulation process for redundant compared to unisensory conditions. This increased noise may come, for instance, from a disruptive interaction between parallel decision units (i.e. having both accumulate evidence at the same time decreases their reliability). This second parameter is motivated by findings showing that empirical RT distributions with redundant signals are not faster but more variable than the best-fitting race model with only Rho as a free parameter¹³. A key advantage of considering entire RT distributions is that these provide a more detailed view compared to an approach that is based only on a subset of particularly fast RTs as in Miller's test. Such a detailed view is critically needed given that present understanding of multisensory interactions is still far from complete. For example, the source of additional noise is not yet known and subject to speculation¹⁹. Thus, it will be particularly important to identify experimental manipulations that specifically target one or the other interaction so that the underlying sources can be determined. In the end, we are convinced that such a comparative approach will lead to a better understanding of multisensory processing.

Following the comparative approach, we developed stimuli to mimic some of the diversity across RSE studies. Using a 2×2 design, we manipulated *stimulus construction* and *signal features* in each modality (Fig. 1b). Regarding the first factor, the construction of auditory and visual stimuli was either simple (identical across trials; without background stimulation) or complex (randomly generated; with background stimulation). One aspect of this manipulation is that simple stimuli had strong onset transients, which were masked in complex conditions by the background. As onset transients have been considered relevant to multisensory processing^{22,23}, we aim to evaluate whether processing interactions, as revealed by the modelling approach, change according to this factor. Regarding the second factor, signal features were either consistent (one variant per modality) or alternating (one of two variants per modality). With this factor, we aimed to understand the origins of history effects. One possibility is that history effects arise simply because of the repetition of low-level features (e.g. the specific frequency of a tone). If true, introducing alternating signals within each modality would reduce history effects. Alternatively, if no changes in history effects occur, this would suggest a basis in higher-level processes.

Methods

Participants. 20 healthy adults (18–29 years; 14 females) were recruited via the University of St Andrews. All were naïve regarding the experiment's purposes and reported normal hearing and normal/corrected-to-normal vision. Informed consent was obtained from all participants. Participants were reimbursed with £9. All procedures were approved by the University of St. Andrews' *University Teaching and Research Ethics Committee* (UTREC, approval code: PS12181) and were performed in accordance with the Code of Human Research Ethics (British Psychological Society, 2014).

Apparatus. The experiment was controlled by a Dell computer (Optiplex XE2) equipped with Matlab and the Psychophysics Toolbox extensions^{24–26}. Visual stimuli were presented on a Dell UltraSharp U2713HM monitor (resolution: $1,920 \times 1,080$ pixel; refresh: 60 Hz). Viewing distance was 57 cm. Auditory stimuli were presented via over-ear headphones (Sennheiser HD 280 Pro) at a 44.1 kHz sample frequency. The volume was calibrated using a Brüel and Kjær sound level meter (Type 2250) attached to an artificial ear (Type 4153). Participants responded by pressing a custom-built handheld button connected to an RTbox V5²⁷. Prior to data collection, the RTbox was also used to calibrate audio-visual timing. As a test, audio-visual signal onsets were presented in 1000 trials. Calibrated stimuli were synchronous and onset jitter was reliably below 1 ms.

Task. We used the redundant signals paradigm (Fig. 1a). On signal trials (75%), we presented either an auditory (A), a visual (V), or redundant signals (both auditory and visual together, AV). Participants pressed a button as quickly as possible after detecting any signal. On catch trials (25%), no signal was presented and participants were asked to withhold a response.

Stimuli. *Auditory Stimuli.* We manipulated *stimulus construction* and *signal features* following a 2×2 design (Fig. 1b). In simple conditions, signals were pure tones. In the consistent condition, the signal was always a 440 Hz tone. In the alternating condition, the signal was either a 440 or 660 Hz tone. Tones were presented at 45 dB SPL. In complex conditions, signals were noise sounds. To generate a new sound on each trial, Gaussian noise (i.e., a sequence of normally-distributed random numbers) was filtered with a 2nd order Butterworth bandpass filter. In the consistent condition, we always used edge frequencies of 1.0/1.1 kHz. In the alternating condition, we used edge frequencies of either 1.0/1.1 kHz or 1.1/1.2 kHz. Noise sounds were presented at 45 dB SPL. Noise sounds were presented within background noise (1st order Butterworth bandpass filter, edge frequencies: 0.5/2.4 kHz, 50 dB SPL). Auditory signals had a ramp onset of 10 ms.

Visual Stimuli. Signals covered the area of a notional annulus with an inner/outer radius of $1^\circ/4^\circ$ (visual angle) around central fixation. In simple conditions, signals were composed of 3 concentric rings (1° increments, alternating black-to-white; Fig. 1b). In the consistent condition, the ring pattern always started on black. In the alternating condition, the ring pattern started either on black or white. In complex conditions, 1000 dots (1 pixel; black or white) were uniformly-distributed within the annulus as visual background stimulation. Each dot moved

linearly in a random direction and speed (mean: 1°/s, SD: 0.2°/s). Dots were randomly replaced within the annulus when moving outside the annulus, or when the lifetime of 0.1 s had passed. For the signal, 50% of the dots started coherent rotation (mean: 0.67 rad/s, SD: 0.067 rad/s). In the consistent condition, the direction was always clockwise. In the alternating condition, the direction was either clockwise or anticlockwise. All visual stimuli were presented on a grey screen.

Procedures. Each trial started with a central, green fixation point (0.27°). In complex conditions, both the auditory and visual background were presented as well. The foreperiod duration consisted of a fixed component (1 s) and an exponentially-distributed random component (mean: 0.75 s). Then, a signal was presented for a max. duration of 1.5 s. On catch trials, no signal was presented and stimulation as during the foreperiod continued for 1.5 s. After a response, or at the end of the max. duration, the fixation point changed from green to red to indicate that the trial ended. Following false alarms and misses, a feedback screen indicated the mistake for 2 s. Stimulation, including the background in complex conditions, stopped for 0.25 s before the next trial started.

Stimuli were presented in blocks of 104 trials (26 trials per modality, 26 catch trials). The trial sequence was randomised, but always started with an additional dummy trial (presenting redundant signals). After a false alarm or miss, a corresponding trial was repeated, the trial sequence was reshuffled, and stimulation continued with another dummy trial. Dummy trials were not analysed. The experiment included 16 blocks (2 × 2 within-subjects design with 4 blocks per condition). Conditions were intermixed in the randomisation of block order, which was done for each participant using a Latin square. A block lasted approx. 4 min (a small pointer on fixation denoted progress). The entire experiment lasted around 105 min with breaks.

Data analysis. Any response within 1.5 s after signal onset was recorded as valid. Responses were false alarms when given during the foreperiod or at any time during catch trials. Signal trials with no response were misses. Before the main analysis, we performed an outlier correction. For this, we transformed the RTs of each condition into rates (1/RT). We excluded data points that deviated by more than $3 \times 1.4826 \times \text{median absolute deviations (MADs)}$ from the median²⁸, corresponding to 3 SDs if data are normally distributed. Fast outliers occurred on 0.4% (± 0.1) of trials and slow outliers on 0.5% (± 0.1) of trials. A total of 24,720 valid RTs (approx. 103 per condition and participant) remained.

The main analysis focused on the speed-up of RTs with redundant signals compared to the unisensory components, which is the RSE. As defined previously, this multisensory benefit is best measured by the area between the cumulative distribution function (CDF) in the multisensory condition and the faster of the unisensory CDFs¹². As a simple procedure, if the number of measured RTs is the same in all three conditions, the size of the area can be estimated:

$$\text{benefit} = \frac{\sum \min(A_i, V_i) - AV_i}{N} \quad (1)$$

A_i , V_i , and AV_i are quantiles of the empirical CDFs, where the index i indicates the rank ranging from 1 (the fastest) to N (the slowest RT).

This simple computation requires equal numbers of RTs in each condition to allow for equal quantile points. However, following outlier correction, the numbers of RTs are often unequal. As a solution, we used linear interpolation²⁹ to down-sample the collected RTs to a common sample size (we used 50 quantile points; the procedure was shown to be unbiased using Monte Carlo simulations). We computed empirical benefits for each participant and each condition of the 2 × 2 design.

To obtain quantitative predictions of the RSE, we used Raab's model¹⁸. Based on the CDFs in the unisensory conditions, a parameter-free prediction of the CDF in the multisensory condition can be obtained using probability summation:

$$P_{AV}(t) = P_A(t) + P_V(t) - P_A(t) \times P_V(t) \quad (2)$$

P_A and P_V are the empirical CDFs in the unisensory conditions (we used linear interpolation to obtain continuous CDFs). P_{AV} is here the predicted CDF for multisensory condition. We then extracted 50 quantile RTs from the predicted CDF and calculated expected benefits analogous to the calculation of empirical benefits (Equation (1)).

Next, we quantified two processing interactions that go beyond Raab's model. First, RTs on a given trial can depend on previously presented signals. To quantify this history effect, we computed the RT difference between trials in which modalities were repeated, and trials in which modalities were switched:

$$\text{history effect} = \overline{RT}_{\text{Switch}} - \overline{RT}_{\text{Repetition}} \quad (3)$$

We computed the history effect for auditory and visual RTs separately and averaged these values to obtain a single measure. Second, we computed violations of Miller's bound⁸, which is given by the sum of the CDFs in the unisensory conditions:

$$P_{\text{Miller}}(t) = \min[P_A(t) + P_V(t), 1] \quad (4)$$

The CDF in the redundant signals condition has to fall below Miller's bound if both the race model and context invariance hold¹⁹. To quantify violations, we first used linear interpolation to obtain 50 quantiles for Miller's bound (Miller_i). Then, similar to a previously detailed method³⁰, we computed the size of the area between the CDF in the redundant condition and Miller's bound, where the redundant condition exceeded the bound:

$$violation = \frac{\sum \max(Miller_i - AV_i, 0)}{N} \quad (5)$$

AV_i are the 50 quantile RTs from the empirical CDF in the redundant condition as used in the computation of benefits.

Finally, we applied the context variant race model^{12,13}. All modelling was performed in the rate space (1/RT). As a first step, we fitted a LATER model^{31,32} to each of the unisensory conditions. As the LATER model assumes that rates are normally distributed, we used the Matlab function *normfit* to obtain best-fitting values of the mean (*Mu*) and the standard deviation (*Sigma*) using the Minimum-Variance Unbiased Estimator (MVUE). As a second step, we fitted the context variant race model in the redundant condition. Using a race mechanism, the model assumes that a response on a given trial with redundant signals is triggered by the unisensory signal with the higher rate. Hence, the resulting rate distribution can be computed using the maximum distribution of two Gaussian random numbers³³, which are given by the LATER fits in the unisensory conditions. The exact distribution depends on the correlation (*Rho*), which we kept as a free parameter to account for history effects. As a second free parameter, the additional noise (*Eta*) increases the variability of rates (i.e. the *Sigma* parameter) in both LATER units additively. This free parameter manifests a violation of the context invariance assumption, which allows for violations of Miller's bound. We used the Matlab function *mle* to obtain maximum likelihood estimates (MLEs) of *Rho* and *Eta* individually for each participant and each condition of the 2×2 design.

Statistical analyses were performed in IBM SPSS Statistics 22. For main analyses, we used 2×2 repeated-measures ANOVAs, which always tested the factors stimulus construction (simple, complex) and signal features (consistent, alternating). For initial analysis of median RTs, we used $2 \times 2 \times 3$ repeated-measures ANOVAs, including the additional factor signal modality (A, V, and AV). We used a Greenhouse-Geisser correction where sphericity was violated. The alpha level for statistical testing was 0.05.

Results

General Performance. To assess data quality, we first checked performance (false alarms, misses). In signal trials, the false alarm rate during the foreperiod was 1.04% ($\pm 0.19\%$, SEM). In catch trials, the false alarm rate was 1.53% ($\pm 0.25\%$). The miss rate was 0.46% ($\pm 0.19\%$). As performance was close to perfect, it is not further considered in the RT analysis (for a detailed analysis of false alarm and miss rates, see Supplementary Analysis 1).

To assess whether the experimental manipulations mimicked the diversity of RSE studies, we next tested median RTs (Table S1). A $2 \times 2 \times 3$ repeated-measures ANOVA showed a significant main effect of stimulus construction, $F(1, 19) = 154.59$, $p < 0.001$, $\eta^2 = 0.89$. RTs were faster with simple (0.300 ± 0.010 s) than with complex stimuli (0.416 ± 0.015 s). Hence, the experimental factor stimulus construction successfully manipulated RTs. There was neither a main effect of signal features nor any interaction (all $F \leq 2.04$, $p \geq 0.165$, $\eta^2 \leq 0.10$). Expectedly, there was a main effect of signal modality, $F(1.39, 26.37) = 53.32$, $p < 0.001$, $\eta^2 = 0.74$. Pairwise comparisons revealed that redundant RTs (0.321 ± 0.011 s) were faster than both auditory (0.371 ± 0.014 s) and visual RTs (0.382 ± 0.012 s), both $p < 0.001$. Auditory and visual RTs were not significantly different, $p = 0.626$. This multisensory benefit is the RSE, which we investigate in full detail using our comparative approach.

Additionally, we also examined effects on variability of RTs, as measured by the Median Absolute Deviation (MAD) of RTs. A $2 \times 2 \times 3$ repeated-measures ANOVA showed a main effect of stimulus construction, $F(1, 19) = 58.851$, $p < 0.001$, $\eta^2 = 0.76$. MAD of RTs was smaller with simple (0.040 ± 0.003 s) compared to complex (0.060 ± 0.004 s) stimuli. There was also a main effect of signal modality, $F(1, 19) = 35.306$, $p < 0.001$, $\eta^2 = 0.65$. As expected from probability summation, redundant RTs were less variable (0.040 ± 0.003 s) than auditory (0.057 ± 0.004 s) and visual (0.054 ± 0.004 s) RTs, both $p < 0.001$. MAD of auditory and visual RTs was not significantly different, $p = 0.822$. Finally, there was a significant interaction between stimulus construction and signal modality, $F(2, 38) = 5.230$, $p < 0.010$, $\eta^2 = 0.22$. Paired-samples were conducted for MAD (pooled across signal features) for each modality. This analysis found larger MAD values for complex conditions compared to simple conditions for all three modalities, all $p < 0.001$. The interaction is explained by the difference between simple and complex MAD values being larger for auditory (0.021 s) and visual (0.024 s) signals than it is for redundant signals (0.013 s).

Step 1: Multisensory Benefits. As first step, we tested the ability of the race model framework to predict multisensory benefits, which are best measured based on RT distributions (Equation (1); for an illustration, see Fig. 2a). According to Raab's model, a parameter-free prediction of the RT distribution with redundant signals can be obtained using probability summation (Equation (2)), which allows calculation of predicted benefits analogous to empirical benefits. For predicted benefits (Supplementary Fig. S1), a 2×2 repeated-measures ANOVA showed a main effect of stimulus construction, $F(1, 19) = 41.54$, $p < 0.001$, $\eta^2 = 0.69$. Predicted benefits were larger for complex (0.046 ± 0.004 s) compared to simple stimuli (0.032 ± 0.003 s). No other effects were significant (all $F \leq 0.80$, $p \geq 0.383$, $\eta^2 \leq 0.04$). Hence, this simple analysis based on the unisensory RTs predicts larger benefits for complex compared to simple stimuli (which is here basically driven by the variability rule, see Supplementary Analysis 2).

We tested next if the RSE followed the prediction. For empirical benefits (Supplementary Fig. S1), a 2×2 repeated-measures ANOVA showed a main effect of stimulus construction, $F(1, 19) = 5.87$, $p < 0.026$, $\eta^2 = 0.24$. Empirical benefits were larger for complex (0.052 ± 0.003 s) compared to simple stimuli (0.041 ± 0.004 s). No other effects were significant (all $F \leq 0.41$, $p \geq 0.528$, $\eta^2 \leq 0.02$). Hence, empirical benefits were consistent with the prediction.

To assess the explanatory power, we correlated predicted and empirical benefits (Fig. 2b). All correlation coefficients were positive, and significant in 3 of 4 conditions (Table S2). One notable observation is that mean

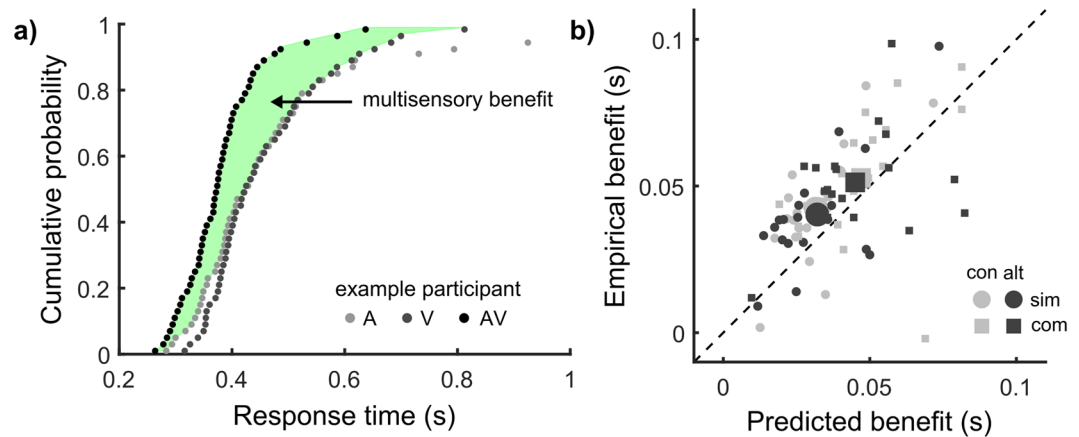


Figure 2. Analysing multisensory benefits. (a) Empirical benefits are measured by the area between the cumulative RT distributions in the multisensory (AV) condition and the faster of the unisensory conditions (A, V; Equation (1)). Data from an example participant in the complex-alternating condition. (b) Empirical benefits as a function of benefits predicted by Raab's model (Equation (2)). Each point represents a participant in one of the four conditions. Large symbols represent the group mean.

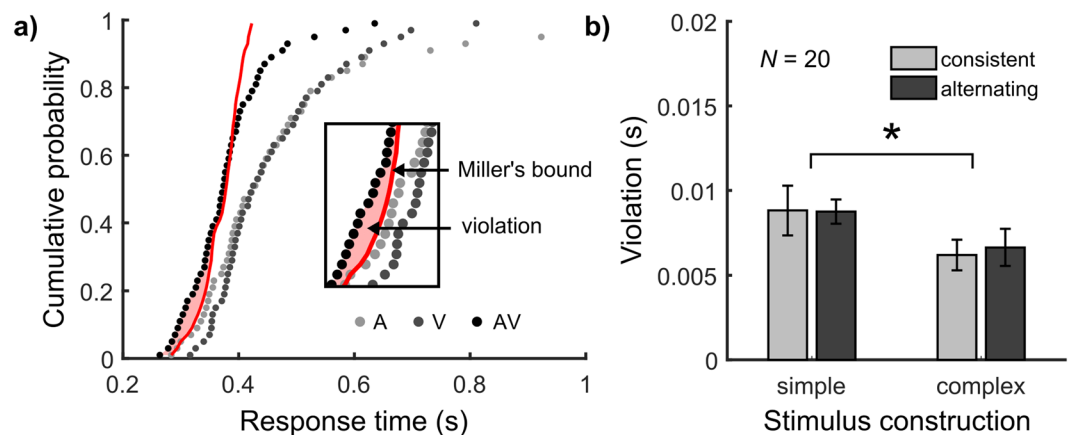


Figure 3. Measuring empirical interactions. (a) Violations of Miller's bound are measured by the area between the cumulative RT distributions in the multisensory (AV) and the sum of the unisensory conditions (Equations (4) and (5)). Example data as in Fig. 2a. (b) Violations of Miller's bound across conditions. Mean and SEM of 20 participants.

empirical benefits for each participant over all conditions (0.047 ± 0.003 s) was larger than mean predicted benefits (0.039 ± 0.003 s). This was confirmed by a paired-samples *t*-test, $t(19) = 3.07$, $p = 0.006$. Thus, while Raab's parameter-free model broadly predicts the RSE, it does not provide a complete account.

Step 2: Empirical Interactions. As second step, we focus on two processing interactions that are not covered by Raab's model. Firstly, we measured history effects by analysing RTs as a function of previously presented signals (Equation (3)). Expectedly, a 2×2 repeated-measures ANOVA showed a significant intercept of 0.034 s (± 0.004 s), $F(1, 19) = 75.62$, $p < 0.001$, $\eta^2 = 0.80$ (Fig. S3). Hence, unisensory RTs following a modality switch were slower compared to a repetition. If history effects arise simply due to the repetition of low-level signal features (e.g. the specific frequency of a tone), introducing alternating features within each modality should reduce history effects compared to consistent features. However, no effects were significant (all $F \leq 4.13$, $p \geq 0.056$, $\eta^2 \leq 0.179$). Hence, while clearly present and thus to be considered by any model of the RSE, the history effect did not change across conditions here (for additional tests, see Supplementary Analysis 3).

Secondly, we measured violations of Miller's bound (Equations (4) and (5); for an illustration, see Fig. 3a). A 2×2 repeated-measures ANOVA showed a significant intercept of 0.008 s (± 0.001 s), $F(1, 19) = 130.70$, $p < 0.001$, $\eta^2 = 0.87$ (Fig. 3b). Hence, in agreement with most RSE studies using audio-visual signals, Miller's bound was violated. In addition, there was a significant effect of stimulus construction $F(1, 19) = 4.56$, $p < 0.046$, $\eta^2 = 0.19$. The violation area for simple stimuli (0.009 ± 0.001 s) was larger than for complex stimuli (0.006 ± 0.001 s). No other effects were significant (all $F \leq 0.05$, $p \geq 0.819$, $\eta^2 < 0.01$). Thus, violations are clearly present and therefore must be considered by race models.

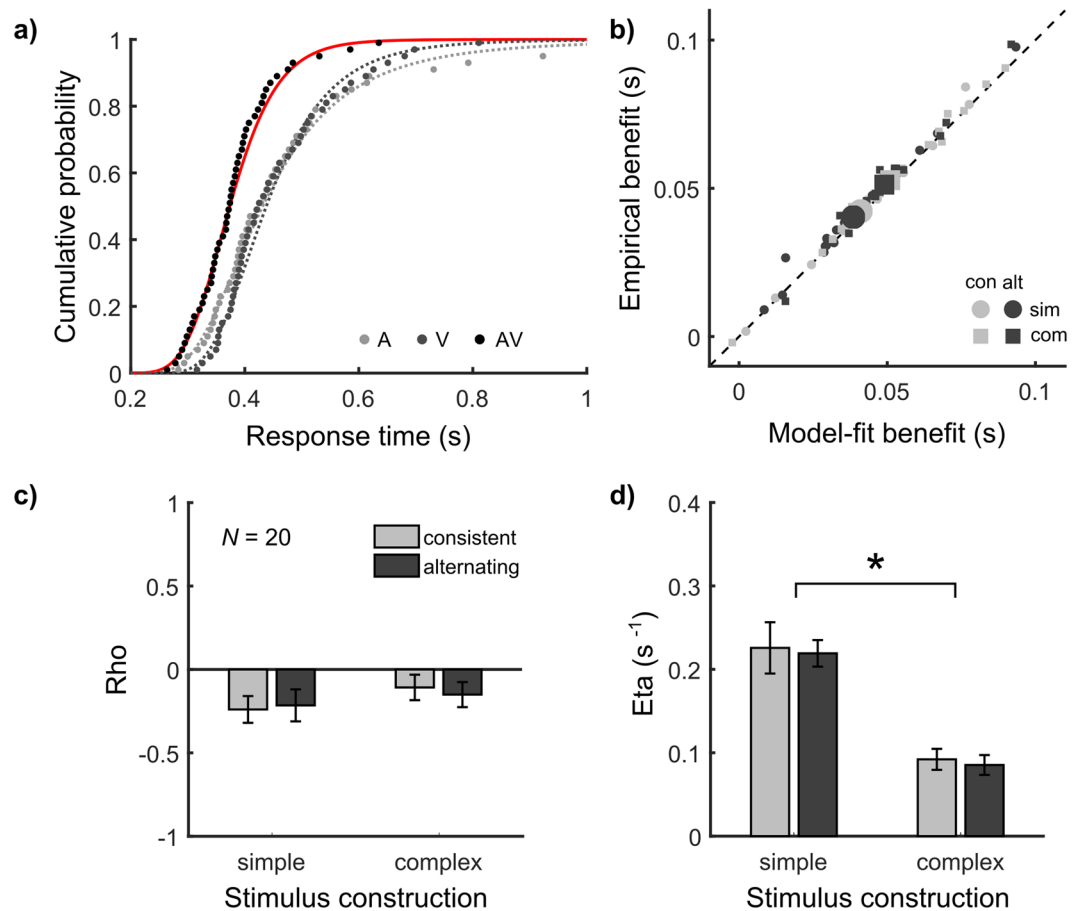


Figure 4. Model fitting. (a) Best-fitting context variant race model (solid line). The model is constrained by the fits of the LATER model in the unisensory conditions (dashed lines) and has the correlation Rho and the additional noise η as free parameters. Example data as in Fig. 2a. (b) Empirical benefits as a function of benefits calculated from the best-fitting context variant race model. Each point represents a participant in one of the four conditions. Large symbols represent the group mean. (c,d) Best fitting values of Rho and η across conditions. Mean and SEM of 20 participants.

Step 3: Modelling Approach. As final step, we followed a modelling approach that accounts for the empirical interactions. We used the context variant race model^{12,13}, which includes the correlation Rho and the noise η as free parameters to account for history effects and violations of Miller's bound, respectively. We fitted the model on the level of RT distributions (for an illustration, see Fig. 4a). To assess if the best-fitting model fully explained the RSE, we computed benefits according to the model-fit analogous to empirical benefits (Equation (1)). As with the parameter-free predictions, we correlated model-fit and empirical benefits (Fig. 4b). The model explained the empirical benefits almost perfectly as all correlation coefficients were positive and highly significant (see Supplementary Table S2). Hence, the context variant race model provides a viable explanation for the RSE across conditions.

We next inspected the best-fitting model parameters with respect to the empirical interactions. Regarding the first parameter Rho , a 2×2 repeated-measures ANOVA showed a significant intercept of $-0.178 (\pm 0.064)$, $F(1, 19) = 7.72$, $p = 0.012$, $\eta^2 = 0.29$ (Fig. 4c). Hence, there was overall a negative correlation between unisensory RTs. There were no further significant effects (all $F \leq 1.24$, $p \geq 0.279$, $\eta^2 \leq 0.06$). This finding is in agreement with the above finding that history effects did not differ across conditions. Regarding the second parameter η , a 2×2 repeated-measures ANOVA showed a significant intercept of $0.156 s^{-1} (\pm 0.012 s^{-1})$, $F(1, 19) = 176.06$, $p < 0.001$, $\eta^2 = 0.90$ (Fig. 4d). The parameter manifests a violation of the context invariance assumption, which allows for violations of Miller's bound. We neither found a main effect of signal features nor an interaction (all $F \leq 0.15$, $p \geq 0.708$, $\eta^2 \leq 0.01$). Interestingly, there was a main effect of stimulus construction, $F(1, 19) = 44.62$, $p < 0.001$, $\eta^2 = 0.70$. η was larger with simple ($0.222 \pm 0.019 s^{-1}$) compared to complex stimuli ($0.089 \pm 0.010 s^{-1}$). In agreement with the above finding that the violation of Miller's bound is larger for simple compared to complex stimuli, more noise was required to account for the redundant RT distributions with simple compared to complex stimuli according to the model. The analysis of η thus points to a striking dissociation between the processing interactions revealed here (Fig. 4d) and multisensory benefits, which in contrast were larger with complex compared to simple signals (Supplementary Fig. S1).

Discussion

To mimic the diversity of RSE studies, we used a factorial design to create distinct audio-visual stimulus sets. With our sets, RTs were faster for simple compared to complex stimuli (with a rather substantial RT difference of 0.116 s). The comparative approach outlined in this paper, therefore, can be considered a blueprint to systematically analyse the RSE with distinct stimulus sets or different participant groups, which we think is critically needed to gain complete understanding of the RSE.

As a first step to this approach, we focused on multisensory benefits as given by the size of the RSE. Initially, we made a quantitative prediction of the RSE based on Raab's model¹⁸. At this point, we would like to highlight that this basic race model is fully constrained by the empirical RT distributions in the unisensory conditions, and does not contain free parameters. The quantitative prediction is simply computed based on probability summation (Equation (2)). Despite its simplicity, the model correctly predicted the increased RSE as observed for complex rather than simple stimuli (Fig. 2) and accounts considerably for the variation across participants (Table S2). To the best of our knowledge, no such straightforward predictions are provided by any published pooling model, which makes it difficult to understand why this alternative model class is favoured in RSE research e.g.³⁴. In contrast, we argue that race models are a very promising candidate for a common framework for the diversity of RSE studies. At least, Raab's model provides a simple tool for analysing the RSE when distinct signal sets or different participant groups yield different RTs.

As second and third step, we analysed two empirical interactions not accounted for by Raab's model, and applied consequently the context variant race model, which is an extended race model including two free parameters to cover the interactions^{12,13}. The first interaction concerns history effects, which challenges statistical independence (as assumed in Equation (2)). In agreement with previous studies^{4,8,13}, we found significant history effects across conditions, which thus must be accounted for by any model of the RSE. In the race model, we dropped the statistical independence assumption and included instead the correlation *Rho* as a free parameter. As history effects are opposite for vision and audition (e.g. after presentation of a visual signal on the previous trial, fast visual and slow auditory RTs are expected), the best-fitting *Rho* value covered the effect by assuming a negative correlation (Fig. 4c).

One simple hypothesis, which we began with, is that history effects may arise due to the repetition of low-level signal features. For example, history effects may arise in auditory trials because tones with identical frequencies repeatedly stimulate the same population of sensory neurons. If this repeated stimulation is causally-linked, the history effect should be reduced when tones with different frequencies stimulate different populations of sensory neurons. Following this simple explanation, we expected to observe a main effect of signal features, whereby alternating signals would reduce history effects compared to consistent signals. However, such changes occurred neither for the measured history effect nor for the corresponding model parameter *Rho*. This suggests rather a basis in higher-level processes. For instance, in the field of task-switching³⁵, costs are observed by switching between two tasks in the same modality. History effects may be the result of similar mechanisms (i.e. switching between auditory and visual detection tasks). Alternatively, in the attention literature, RT costs have been observed between targets in different modalities³⁶. History effects may also arise from attentional mechanisms, whereby the previous unisensory trial directs attention to one modality at the expense of the other. Future investigations should develop manipulations to target such higher-level processes.

The second interaction concerns violations of Miller's Bound, which challenges the context invariance assumption¹⁹. In agreement with most audio-visual RSE studies e.g.^{4,5,8,9,13}, we found significant violations of Miller's bound across conditions (Fig. 3b), which thus must be accounted for by race models. In the context variant race model^{12,13}, we dropped the context invariance assumption and included instead the additional noise *Eta* as a second free parameter. Critically, for violations of Miller's bound we found a main effect of stimulus construction, with increased violations occurring with simple compared to complex stimuli. The best-fitting *Eta* value covered the effect by assuming further increased noise in the evidence accumulation process for simple compared to complex conditions (Fig. 4d).

Previously, we have only speculated about potential sources of this additional noise¹⁹. Here, however, one aspect of the experimental factor of stimulus construction is that simple stimuli are expected to trigger strong onset transients, which are masked by the background stimulation in complex conditions. As onset transients have been considered relevant for multisensory processing^{22,23}, one possibility is that the additional noise interaction is linked to an interference caused by these transients. Although our behavioural data does not allow us to identify exact sources, one interpretation could be that strong neuronal activity linked to sudden signal onsets leaks as noise across sensory modalities and otherwise separate parallel processes.

As a more general outcome of our comparative approach, the systematic analysis of the RSE calls for a clarification of the rather vague term "multisensory integration". On the one hand, the term has been defined operationally as "a multisensory response (neural or behavioural) that is significantly different from the responses evoked by the modality-specific component stimuli"³⁷. Following this definition, we found that the multisensory benefit, and hence integration, was larger with complex compared to simple signals. On the other hand, in research on multisensory RTs, multisensory integration is often said to occur only if Miller's bound is violated (which is even considered "a psychometric benchmark of integrative processing"³⁸). Regarding this definition, we found the opposite. Violations of Miller's bound (and consistently the best-fitting *Eta* value) were larger with simple compared to complex signals. Hence, the two variables clearly dissociate and should not be described using the same term or share the same concept.

The vagueness of definitions may have contributed to the lack of a common framework for the RSE across studies. In contrast, our comparative approach, which is based on Raab's original race model¹⁸, provides conceptual clarity. The basic model architecture, which consists of parallel decision units coupled by a logic OR operator, is convincing as it perfectly matches the task demands as imposed by the redundant signals paradigm¹⁹. Moreover, it provides a strong, parameter-free predictor of multisensory benefits¹². As such, the combination rule

for redundantly defined signals (or the multisensory integration mechanism, under an alternative conception) is clearly defined. Beyond this basic mechanism, at least two very specific processing interactions emerge and can be precisely quantified by explicit parameters in a model-based approach.

The approach is general enough to be applied to various experimental setups and research questions. This includes not only experiments comparing different sensory modalities, such as audio-tactile⁴ or visual-tactile stimulation⁵, but also studies that investigate unisensory RSE^{39,40} or processing differences between uni- and multisensory versions of the redundant signals paradigms⁴¹. The approach is also suitable to shed light on processing differences between participant groups. For instance, studies on ageing typically show larger multisensory benefits for older compared to younger participants⁹. Such findings may be directly understood if older participants' RT distributions are more variable, as unisensory RT variability is the driving force behind multisensory benefits according to race models¹². Finally, since the framework has successfully been used to study dyslexia⁴², it could of course also be applied to study other clinical populations, such as those with Schizophrenia⁴³ or Parkinson's Disease⁴⁴. In the aggregate, we believe that our comparative approach provides both a general account and the precision needed to target and understand processing differences across diverse studies, which are key features for a common explicatory framework of the RSE.

Beyond the RSE, we believe this work contributes to a broader effort to achieve clarity of definitions in multisensory research³⁷. In the wider literature, there are many instances of behavioural paradigms and associated measures assumed to reflect 'multisensory integration' equivalently. As shown here within one paradigm, however, this assumption can be incorrect. We hope therefore to demonstrate for multisensory researchers the importance of defining measures in detail and verifying their proposed relationships experimentally. Failing to do so may result in similar confusion when comparing results between paradigms.

Data Availability

The research data supporting this publication can be accessed at <https://doi.org/10.17630/c8cbd7b7-e2b3-4e62-bb3d-66fce081ff59>⁴⁵. The data analysis and model-fitting tools are also freely available online via the RSE-box⁴⁶: <https://github.com/tomotto/RSE-box>.

References

1. Hershenson, M. Reaction-Time as a Measure of Intersensory Facilitation. *J Exp Psychol* **63**, 289–&, <https://doi.org/10.1037/h0039516> (1962).
2. Kinchla, R. A. Detecting Target Elements in Multielement Arrays - Confusability Model. *Percept Psychophys* **15**, 149–158, <https://doi.org/10.3758/Bf03205843> (1974).
3. Todd, J. W. Reaction to multiple stimuli. *Arch. Psychol.*, 1–65 (1912).
4. Gondan, M., Lange, K., Rosler, F. & Roder, B. The redundant target effect is affected by modality switch costs. *Psychon B Rev* **11**, 307–313, <https://doi.org/10.3758/Bf03196575> (2004).
5. Forster, B., Cavina-Pratesi, C., Aglioti, S. M. & Berlucchi, G. Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Exp Brain Res* **143**, 480–487, <https://doi.org/10.1007/s00221-002-1017-9> (2002).
6. Veldhuizen, M. G., Shepard, T. G., Wang, M. F. & Marks, L. E. Coactivation of Gustatory and Olfactory Signals in Flavor Perception. *Chem Senses* **35**, 121–133, <https://doi.org/10.1093/chemse/bjp089> (2010).
7. Bolognini, N., Olgiati, E., Rossetti, A. & Maravita, A. Enhancing multisensory spatial orienting by brain polarization of the parietal cortex. *Eur J Neurosci* **31**, 1800–1806, <https://doi.org/10.1111/j.1460-9568.2010.07211.x> (2010).
8. Miller, J. Divided Attention - Evidence for Co-Activation with Redundant Signals. *Cognitive Psychol* **14**, 247–279, [https://doi.org/10.1016/0010-0285\(82\)90010-X](https://doi.org/10.1016/0010-0285(82)90010-X) (1982).
9. Peiffer, A. M., Mozolic, J. L., Hugenschmidt, C. E. & Laurienti, P. J. Age-related multisensory enhancement in a simple audiovisual detection task. *Neuroreport* **18**, 1077–1081, <https://doi.org/10.1097/WNR.0b013e3281e72ae7> (2007).
10. Molholm, S., Ritter, W., Javitt, D. C. & Foxe, J. J. Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cereb Cortex* **14**, 452–465, <https://doi.org/10.1093/cercor/bhh007> (2004).
11. Collignon, O. *et al.* Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia* **48**, 220–225, <https://doi.org/10.1016/j.neuropsychologia.2009.09.007> (2010).
12. Otto, T. U., Dassy, B. & Mamassian, P. Principles of Multisensory Behavior. *J Neurosci* **33**, 7463–7474, <https://doi.org/10.1523/Jneurosci.4678-12.2013> (2013).
13. Otto, T. U. & Mamassian, P. Noise and Correlations in Parallel Perceptual Decision Making. *Curr Biol* **22**, 1391–1396, <https://doi.org/10.1016/j.cub.2012.05.031> (2012).
14. Gondan, M. & Minakata, K. A tutorial on testing the race model inequality. *Atten Percept Psycho* **78**, 723–735, <https://doi.org/10.3758/s13414-015-1018-y> (2016).
15. Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu Rev Neurosci* **30**, 535–574, <https://doi.org/10.1146/annurev.neuro.29.051605.113038> (2007).
16. Bogacz, R. Optimal decision-making theories: linking neurobiology with behaviour. *Trends Cogn Sci* **11**, 118–125, <https://doi.org/10.1016/j.tics.2006.12.006> (2007).
17. Shadlen, M. N. & Kiani, R. Decision Making as a Window on Cognition. *Neuron* **80**, 791–806, <https://doi.org/10.1016/j.neuron.2013.10.047> (2013).
18. Raab, D. H. Statistical Facilitation of Simple ReactionTimes. *T New York Acad Sci* **24**, 574–590, <https://doi.org/10.1111/j.2164-0947.1962.tb01433.x> (1962).
19. Otto, T. U. & Mamassian, P. Multisensory Decisions: the Test of a Race Model, Its Logic, and Power. *Multisens Res* **30**, 1–24, <https://doi.org/10.1163/22134808-00002541> (2017).
20. Ashby, F. G. & Townsend, J. T. Varieties of Perceptual Independence. *Psychol Rev* **93**, 154–179, <https://doi.org/10.1037/0033-295x.93.2.154> (1986).
21. Luce, R. D. *Response Times: Their Role in Inferring Elementary Mental Organization*. (Oxford University Press 1986).
22. Jaekl, P., Perez-Bellido, A. & Soto-Faraco, S. On the 'visual' in 'Audio-visual integration': a hypothesis concerning visual pathways. *Exp Brain Res* **232**, 1631–1638, <https://doi.org/10.1007/s00221-014-3927-8> (2014).
23. Werner, S. & Noppeney, U. The contributions of transient and sustained response codes to audiovisual integration. *Cereb Cortex* **21**, 920–931, <https://doi.org/10.1093/cercor/bhq161> (2011).
24. Brainard, D. H. The psychophysics toolbox. *Spatial Vision* **10**, 433–436, <https://doi.org/10.1163/156856897x00357> (1997).
25. Kleiner, M., Brainard, D. & Pelli, D. What's new in Psychtoolbox-3? *Perception* **36**, 14–14 (2007).
26. Pelli, D. G. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* **10**, 437–442, <https://doi.org/10.1163/156856897x00366> (1997).

27. Li, X. R., Liang, Z., Kleiner, M. & Lu, Z. L. RTbox: A device for highly accurate response time measurements. *Behav Res Methods* **42**, 212–225, <https://doi.org/10.3758/Brm.42.1.212> (2010).
28. Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J Exp Soc Psychol* **49**, 764–766, <https://doi.org/10.1016/j.jesp.2013.03.013> (2013).
29. Ulrich, R., Miller, J. & Schroter, H. Testing the race model inequality: An algorithm and computer programs. *Behav Res Methods* **39**, 291–302, <https://doi.org/10.3758/Bf03193160> (2007).
30. Colonius, H. & Diederich, A. The race model inequality: interpreting a geometric measure of the amount of violation. *Psychol Rev* **113**, 148–154, <https://doi.org/10.1037/0033-295X.113.1.148> (2006).
31. Carpenter, R. H. & Williams, M. L. Neural computation of log likelihood in control of saccadic eye movements. *Nature* **377**, 59–62, <https://doi.org/10.1038/377059a0> (1995).
32. Noorani, I. & Carpenter, R. H. S. The LATER model of reaction time and decision. *Neurosci Biobehav Rev* **64**, 229–251, <https://doi.org/10.1016/j.neubiorev.2016.02.018> (2016).
33. Nadarajah, S. & Kotz, S. Exact distribution of the max/min of two Gaussian random variables. *Ieee T Vlsi Syst* **16**, 210–212, <https://doi.org/10.1109/Tvlsi.2007.912191> (2008).
34. Chandrasekaran, C. Computational principles and models of multisensory integration. *Curr Opin Neurobiol* **43**, 25–34, <https://doi.org/10.1016/j.conb.2016.11.002> (2017).
35. Monsell, S. Task switching. *Trends Cogn Sci* **7**, 134–140, [https://doi.org/10.1016/s1364-6613\(03\)00028-7](https://doi.org/10.1016/s1364-6613(03)00028-7) (2003).
36. Spence, C., Nicholls, M. E. R. & Driver, J. The cost of expecting events in the wrong sensory modality. *Percept Psychophys* **63**, 330–336, <https://doi.org/10.3758/Bf03194473> (2001).
37. Stein, B. E. *et al.* Semantic confusion regarding the development of multisensory integration: a practical solution. *Eur J Neurosci* **31**, 1713–1720, <https://doi.org/10.1111/j.1460-9568.2010.07206.x> (2010).
38. Martuzzi, R. *et al.* Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cereb Cortex* **17**, 1672–1679, <https://doi.org/10.1093/cercor/bhl077> (2007).
39. Corballis, M. C. Hemispheric interactions in simple reaction time. *Neuropsychologia* **40**, 423–434, [https://doi.org/10.1016/S0028-3932\(01\)00097-5](https://doi.org/10.1016/S0028-3932(01)00097-5) (2002).
40. Schroter, H., Ulrich, R. & Miller, J. Effects of redundant auditory stimuli on reaction time. *Psychon B Rev* **14**, 39–44, <https://doi.org/10.3758/Bf03194025> (2007).
41. Girard, S., Pelland, M., Lepore, F. & Collignon, O. Impact of the spatial congruence of redundant targets on within-modal and cross-modal integration. *Exp Brain Res* **224**, 275–285, <https://doi.org/10.1007/s00221-012-3308-0> (2013).
42. Harrar, V. *et al.* Multisensory integration and attention in developmental dyslexia. *Curr Biol* **24**, 531–535, <https://doi.org/10.1016/j.cub.2014.01.029> (2014).
43. Williams, L. E., Light, G. A., Braff, D. L. & Ramachandran, V. S. Reduced multisensory integration in patients with schizophrenia on a target detection task. *Neuropsychologia* **48**, 3128–3136, <https://doi.org/10.1016/j.neuropsychologia.2010.06.028> (2010).
44. Plat, F. M., Praamstra, P. & Horstink, M. W. Redundant-signals effects on reaction time, response force, and movement-related potentials in Parkinson's disease. *Exp Brain Res* **130**, 533–539, <https://doi.org/10.1007/s002219900276> (2000).
45. Innes, B. R. & Otto, T. U. Data underpinning - A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent, <https://doi.org/10.17630/c8cbd7b7-e2b3-4e62-bb3d-66fce081ff59>.
46. Otto, T. U. An analysis and modelling toolbox to study multisensory response times. Abstract from 19th International Multisensory Research Forum (IMRF), online at http://imrf.info/wp_imrf/wp-content/uploads/2018/06/IMRF2018_Abtracts.pdf (2018).

Acknowledgements

We would like to thank Akira O'Connor, Cleo Pike, and Yue (Serena) Liu for their feedback on an earlier draft of the manuscript. This work was supported by the *Biotechnology and Biological Sciences Research Council* (BBSRC, grant number: BB/J01446X/1).

Author Contributions

Both authors designed the experiment. B.R.I. implemented the research, collected and analysed the data, and wrote the initial draft. T.U.O. supervised the research. Both authors wrote the final article.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-39924-6>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019